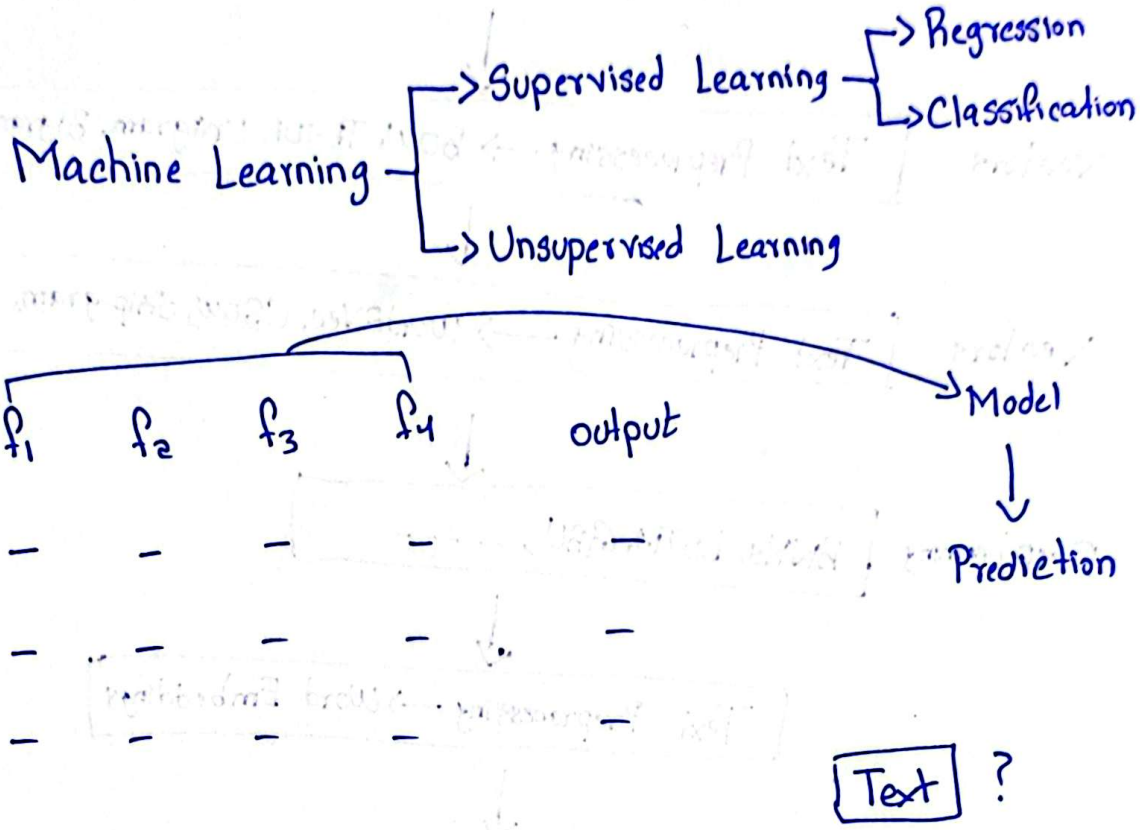
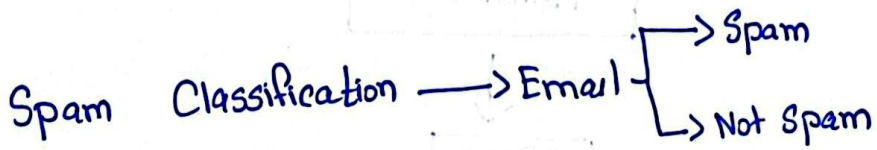
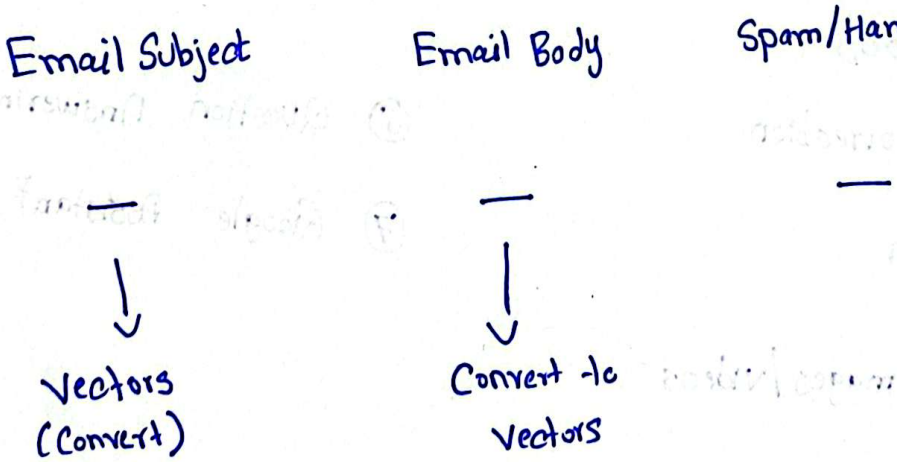
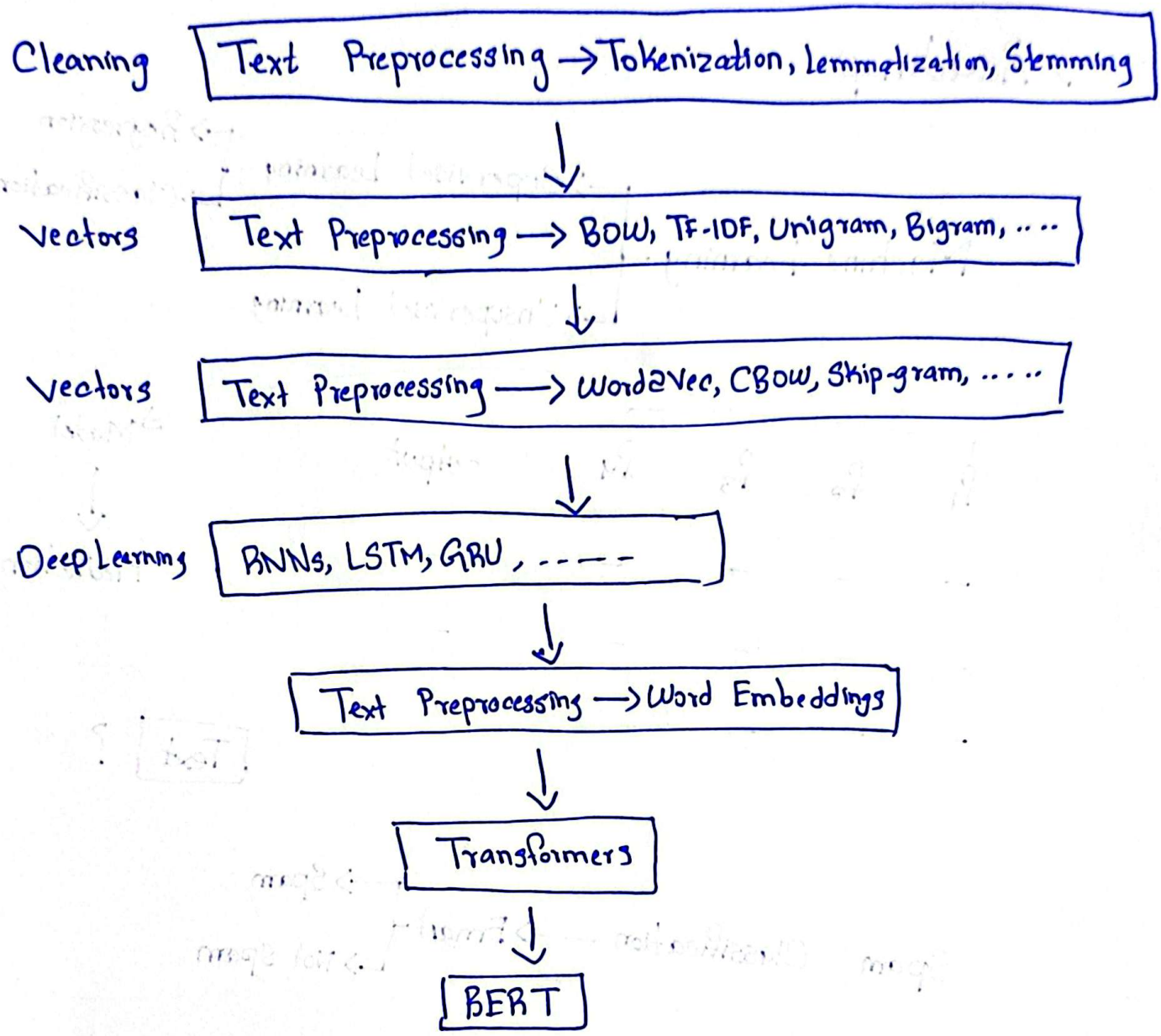
NATURAL LANGUAGE PROCESSING

→ Roadmap









→ Use Cases

1. Spelling correction
2. Translation
3. Text to images/videos
4. Summarization
5. Classification (Text)
6. Question Answering
7. Google Assistant

⟶ Tokenization

Tokenization is a process of converting a paragraph into sentences and sentences into words

① Corpus ⟶ Paragraph

② Documents ⟶ Sentences

③ Vocabulary ⟶ Unique words

④ words ⟶ words

Corpus { "My name is Saad and I have a interest in Machine Learning. I am also a Dade Science Consultant."

⇓

Tokens {sentences}

① My name is Saad and I have a interest in Machine Learning

② I am also a Data Science Consultant

⇓

Tokenization {words}

① My, name, is, Saad, and, I, have, a, interest, in, Machine, Learning

② I, am, also, a, Data, Science, Consultant

⟶ Stemming and Lemmatization

> Stemming

It is the process of removing the last few characters of a given word, to obtain a shorter form, even if that form doesn't have any meaning

> Lemmatization

It is a linguistic process that involves reducing words to their base or root form known as lemma.